

XX Evaluation of safety of a chemical cpd. - using recombinant yeast
 PT expressing human cytochrome p450 and a yeast NADPH-p450 reductase
 XX
 PS Examples: Page 18-21; 124pp; English.
 XX
 CC The amino acid sequence of the human cytochrome p450 species 1A2. The
 CC 1.5 kb cDNA was amplified by PCR using the primers AAQ87733-4. The
 CC product was cloned into the yeast expression vectors pAAH5N or pAHRK to
 CC produce the vectors p1A2 for the expression of the cytochrome p450 alone
 CC or p1A2R co-expressed with the yeast NADPH-p450 reductase, respectively.
 CC The vectors are used in a method for evaluating the safety of a chemical
 CC compound by reacting the chemical compound with recombinantly produced
 CC human cytochrome p450 molecular species 1A2, 2C9 (AAQ87715), 2E1
 CC (AAQ87716) or 3A4 (AAQ87717) or their auxiliary species and variants
 CC (AAQ87718-32) and yeast NADPH-p450 reductase, either as a fused protein
 CC or in cell extracts, and analysing the resulting metabolite to assess the
 CC safety of the chemical compound. The method is useful for determining
 CC whether the chemical compound or its metabolite will be converted into
 CC a carcinogenic or mutagenic form through metabolism in the liver.
 XX
 SQ Sequence 516 AA:

Query Match 99.3%; Score 2625; DB 17; Length 516;
 Best Local Similarity 97.5%; Pred. No. 1.6e-253;
 Matches 50%; Conservative 1; Mismatches 0; Indels 12; Gaps 1;
 QY 1 MALSSQVPSATHELLASAFICLVFWLKGKLPFPVFKGLKSHFPQWGLGRLVLTGKN 60
 DB 1 malssqvpsatellasaifclvtwklprvfkulksppwqywpvlghvltakn 60
 QY 61 PHALSRMSQRYGDVLOIRIGSTPPVLVLSKLDITRQALVRQCDKFGKSPDLYTSLITLS 120
 DB 61 phalsrmsqrygdvlgirgstpvlvlslrlditrgalvrqcdkfgkspdytstltdg 120
 QY 121 QSLTFTSDSPVWAAPRFLAQNALNTFASDPASSSSCYLEHVSKEAKALISRLQELM 180
 DB 121 qslftstdspvwaarrlraqnalntfasiadpassscylehvskeakalislrlqelm 180
 QY 181 AGPGHFDPPYVNVSVANVIGAMCFQHPPESSDEMLSLKNTHEFVETASSGNPLDFPP 240
 DB 181 agphfdppynqvsvanvigamcfqghpessdemslvknthefvetassgnpldfip 240
 QY 241 ILRYLPNAPALQRFKAFNORFIWFLQKTVQHYQDFU-----KNSQZSPASCH 288
 DB 241 ilrylpnapalqrfkafnorfllwflqktvqhyqdfdknsvrditcgaifkbskkgptasgn 300
 QY 289 LIPOEKIVNLVNDIFGAGDTVTVAISWSLMYLVTKPEIQPKIQKELDTVIGREKRPRLS 348
 DB 301 lipqekivnlvndifgagdtvtvaiswslylvtkpeiqkaiqkeldtvigretprls 360
 QY 349 DRPQIPYLEAFLETFRHSSTPPTIPHSTTRDTTLNGCFYIPKCCVFWQWVNVHGFEL 408
 DB 361 drpqipyleafletfrhsstlptphsttrdttdtlngfyipkccvfnvqwnhdpel 420
 QY 409 WEDSEFRPERFLADGTAINKPLSEKXMMPLGKGRKRCIGEVIAKWEIFLFLAILLQGLE 468
 DB 421 wedsefrperfladgtainkplsekmmplgmkgrcigevlakweiflailllqgle 480
 QY 469 FSVPPGVKVDLTPITYGLTMKHARCEHVOARLRFESIN 504
 DB 481 fsvppgvkvdltptygltmkharcehvqarlrfesin 516

RESULT 2

AAQ93167
 ID AAQ93167 standard; Protein: 516 AA.
 XX
 AC AAQ93167;
 XX
 DT 11-OCT-1996 (first entry)
 XX

DE Human cytochrome p450 molecular species 1A2 protein.
 KW Human cytochrome p450; amplified; pckc polymerase chain reaction; primer;
 KW liver; yeast; expression vector; NADPH-p450 reductase; ADH gene promoter;
 KW evaluation; safety; fusion protein; metabolite; detoxification;
 KW carcinogenic.
 XX Homo sapiens.
 PN JP08056695-A.
 XX
 PD 05-MAR-1996.
 XX
 PF 15-JUL-1994; 94JP-0164184.
 XX
 PR 17-JUN-1994; 94JP-0136053.
 PR 20-JUL-1994; 94JP-0201126.
 PR 30-JUL-1994; 94JP-0208279.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 XX
 DR WP1: 1996-182311/19.
 DR N-PSDB: AAT28380.
 XX
 PT Novel method for the evaluation of the safety of a cpd. consist of
 PT human cytochrome p450 and yeast NADPH reductase to determine whether
 PT the analyte cpd. is detoxified or metabolised to a carcinogen.
 XX
 PS Example 1; Page 18-20; 74pp; Japanese.
 CC This is the amino acid sequence of the human cytochrome p450 molecular
 CC species 1A2 protein. The corresp. 1.5 kb fragment encoding the protein
 CC was amplified from a human liver derived cDNA library using primers
 CC AAT26923-4. The prod. was cloned into the yeast expression vectors
 CC pAAH5N to generate plasmid p1A2 for prodn. of the cytochrome only or into
 CC the vector pAHRK to generate the plasmid p1A2R for co-expression with the
 CC yeast NADPH-p450 reductase. The sequence is placed under control of the
 CC yeast ADH gene promoter and terminator.
 CC The vectors are used in a method for evaluating the safety of a cpd. by
 CC reacting the test cpd. with recombinantly produced human cytochrome p450
 CC mol. species 1A2, 2C9 (28381), 2E1 (AAT28482), 3A4 (AAT28483) or their
 CC variants (AAT28384-98) together with yeast NADPH-p450 reductase (either
 CC as a fused protein or as a cell extract) and analysing the resultant
 CC metabolite. The cpd. is considered "safe" if it is not detoxified or
 CC rendered carcinogenic or "unsafe" if it is not detoxified or is
 CC metabolised to a carcinogenic cpd.
 SQ Sequence 516 AA;
 Query Match 99.3%; Score 2625; DB 17; Length 516;
 Best Local Similarity 97.5%; Pred. No. 1.6e-253;
 Matches 50%; Conservative 1; Mismatches 0; Indels 12; Gaps 1;
 QY 1 MALSSQVPSATHELLASAFICLVFWLKGKLPFPVFKGLKSHFPQWGLGRLVLTGKN 60
 DB 1 malssqvpsatellasaifclvtwklprvfkulksppwqywpvlghvltakn 60
 QY 61 PHALSRMSQRYGDVLOIRIGSTPPVLVLSKLDITRQALVRQCDKFGKSPDLYTSLITLS 120
 DB 61 phalsrmsqrygdvlgirgstpvlvlslrlditrgalvrqcdkfgkspdytstltdg 120
 QY 121 QSLTFTSDSPVWAAPRFLAQNALNTFASDPASSSSCYLEHVSKEAKALISRLQELM 180
 DB 121 qslftstdspvwaarrlraqnalntfasiadpassscylehvskeakalislrlqelm 180
 QY 181 AGPGHFDPPYVNVSVANVIGAMCFQHPPESSDEMLSLKNTHEFVETASSGNPLDFPP 240
 DB 181 agphfdppynqvsvanvigamcfqghpessdemslvknthefvetassgnpldfip 240
 QY 241 ILRYLPNAPALQRFKAFNORFIWFLQKTVQHYQDFU-----KNSQZSPASCH 288
 DB 241 ilrylpnapalqrfkafnorfllwflqktvqhyqdfdknsvrditcgaifkbskkgptasgn 300


```

Db 780 CCAACCCGCCGCTCAAGAGCTTTAAGACCTTCAATGATAAAGCTTCGCTGCTGTTCTGCAGA 839
QY 849 aaacagtcacagagagagactacacagactt 876
Db 840 AAACCTCCAGGAGCAGCACTACCAAGACTT 867

RESULT 3
LOCUS BI146179 828 bp mRNA linear EST 05-JUL-2001
DEFINITION 602910487F1 NCL_CGAP_119 Mus musculus cDNA clone IMAGE:5052117 5',
mRNA sequence.
ACCESSION BI146179
VERSION H1146179
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 828)
TITLE NIH-MGC http://mgi.nhl.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11141 row: n column: 22
High quality sequence stop: 793.
FEATURES
Source location/Qualifiers
1..828
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/cclone="IMAGE:5052117"
/cclone_lib="NCL_CGAP_119"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV SPORT6, Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP Library"
BASE COUNT 189 a 233 c 241 g 175 t
ORIGIN

Query Match 27 7%, Score 496.2, DB 10; Length 828,
Best Local Similarity 77 6%, Pref No 1 9a-101;
Matches 640; Conservative 0; Mismatches 178; Indels 7; Gaps 3;

QY 45 tggatcagatgcatgtctccatgtcttctctctctctctctctctctctctctctctctctctct 104
Db 3 TGAATAGATAGAGTTTCTTCAATAT---CTTATAGAGAGAGAGATATAGAGCA 59
QY 105 ctgcatctctctctctctctctctctctctctctctctctctctctctctctctctctctct 164
Db 60 CTCCCATCTCTCTTTAGTCTTCTGATGCTCAGAGCCCTCAAGCAACAGGCTCCCAAG 119
QY 165 gcttgaagatccacacacacacacacacacacacacacacacacacacacacacacacacacac 224
Db 120 GCTGAAGAAATCCACCCGACCCCTGGGGCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 179
QY 225 ggaagacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 284
Db 180 GGAAAGAAATCAATCTTATATATATATATATATATATATATATATATATATATATATATAT 234
QY 285 tctgattggtctctctctctctctctctctctctctctctctctctctctctctctctctctct 344

```

```

Db 240 TCCGATCCGCTCCACTCTCTCTGCTGGTCTGATCAAGGCTTCAACACCAATCAATCAAGTAC 299
QY 345 tgaatcagatgcatgtctccatgtcttctctctctctctctctctctctctctctctctctctct 404
Db 300 TGAATAGATAGAGTTTCTTCAATAT---CTTATAGAGAGAGATATAGAGCA 359
QY 405 ctgcatgacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 464
Db 360 CTAAAGAGAGATGATGATTTTCAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 419
QY 465 gcttgaagatccacacacacacacacacacacacacacacacacacacacacacacacacacac 524
Db 420 GCTGAAGAAATCCACCCGACCCCTGGGGCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 479
QY 525 cctgcatctctctctctctctctctctctctctctctctctctctctctctctctctctctct 584
Db 480 CTTGCTATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
QY 585 agctgagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 644
Db 540 AGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
QY 645 acgcatgagagagagagagagagagagagagagagagagagagagagagagagagagagagag 704
Db 600 AGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
QY 705 gcttgaagatccacacacacacacacacacacacacacacacacacacacacacacacacacac 764
Db 660 ATATGCTGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
QY 765 tcttctctctctctctctctctctctctctctctctctctctctctctctctctctctct 824
Db 720 TCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778
QY 825 agagatctctctctctctctctctctctctctctctctctctctctctctctctctctctct 866
Db 779 ATAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 824

RESULT 4
LOCUS AI529113 987 bp mRNA linear EST 18-MAR-1999
DEFINITION u165d09.y1 Sugano mouse liver mRNA Mus musculus cDNA clone
IMAGE:1887281 5' similar to db:M15053 CYTOCHROME P450 1A2 (HUMAN);
gb:M10522 Mouse cytochrome P-450 gene, complete cds (MOLISE);.
mRNA sequence.
ACCESSION AI529113
VERSION AI529113.1 GI:4443248
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 987)
REFERENCE Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylic, T.,
Underwood, K., Stjepetic, M., Thelshing, H., Allen, M., Howes, Y., Pearson,
B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McAnan, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage@llnl.gov) for further information
MGI:971605
Seq primer: custom primer used
High quality sequence stop: 478.
Location/Qualifiers

```



```

QY 31 cctgctctacagtttggtacagatggcattgtccagctgttccctctctcggccacaga 90
   || || || || || || || || || || || || || || || || || || || || || ||
Db 3 CCCTGATGTTGATAGTATGATAGTATGATAGTATGATAGTATGATAGTATGATAGTAT 54

QY 91 gcttctcctgctctgctcattctctgctggttattctctggtgtctcaggttttgagccc 150
   || || || || || || || || || || || || || || || || || || || || || ||
Db 60 GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119

QY 151 tggggtcccaaaagcctgaaagatccacacacacacacacacacacacacacacacacac 210
   || || || || || || || || || || || || || || || || || || || || || ||
Db 120 CCAGGTTCCTCAAGAGCCTCAAGAGCCTCAAGAGCCTCAAGAGCCTCAAGAGCCTCAAG 174

QY 211 tgtctgacccctgggagagagagagagagagagagagagagagagagagagagagagag 270
   || || || || || || || || || || || || || || || || || || || || || ||
Db 180 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 244

QY 271 ggaagctcagatcagatcagatcagatcagatcagatcagatcagatcagatcagatcag 330
   || || || || || || || || || || || || || || || || || || || || || ||
Db 240 GCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299

QY 331 calccagagagcctcagatcagatcagatcagatcagatcagatcagatcagatcagatc 390
   || || || || || || || || || || || || || || || || || || || || || ||
Db 300 CATCAAGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359

QY 391 ctcacacacacacacacacacacacacacacacacacacacacacacacacacacacac 450
   || || || || || || || || || || || || || || || || || || || || || ||
Db 360 ATTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 419

QY 451 ggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 510
   || || || || || || || || || || || || || || || || || || || || || ||
Db 420 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479

QY 511 agcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 570
   || || || || || || || || || || || || || || || || || || || || || ||
Db 480 GACTCAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 539

QY 571 cagcaggttgagagagatgagagagagagagagagagagagagagagagagagagagag 630
   || || || || || || || || || || || || || || || || || || || || || ||
Db 540 CAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599

QY 631 ggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 690
   || || || || || || || || || || || || || || || || || || || || || ||
Db 600 GCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659

QY 691 ccatgagatgctcagatcagatcagatcagatcagatcagatcagatcagatcagatcag 748
   || || || || || || || || || || || || || || || || || || || || || ||
Db 660 CGAGGAGATGCTGCAACATGCTGCAATTAACAGAGGAGCTTTGCTGAGCAATGCTCA 719

QY 749 ggaagacccctggacttctctcctcctcctcctcctcctcctcctcctcctcctcctc 808
   || || || || || || || || || || || || || || || || || || || || || ||
Db 720 GATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 774

QY 809 ttcagagccttcaacacagaggttctctggttctctgagagagagagagagagagagag 867
   || || || || || || || || || || || || || || || || || || || || || ||
Db 780 TCTAAGAACCTTCACAGGATATCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 839

QY 868 tcaagga 873
   || || || || || || || || || || || || || || || || || || || || || ||
Db 840 ACAAGA 845

```

RESULT 6

```

BI146311
LOCUS 602910860F1 NCI_CGAP_119 Mus musculus cDNA clone IMAGE:5051943 5',
DEFINITION 754 bp mRNA linear EST 05-JUL-2001
mRNA sequence.
ACCESSION BI146311
VERSION BI146311.1 GI:14606312
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,

```

Mammalia, Eutheria, Rodentia, Scurionathii, Muridae, Murinae; Mus.
 1 (bases 1 to 754)
 NIH-MGC <http://mgs.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps@mail.nih.gov
 Tissue procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLAM1141 row: g column: 16
 High quality sequence start: 3
 High quality sequence stop: 746.
 Location/Qualifiers
 1..754

FEATURES

source
 1..754
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:5051943"
 /clone_lib="NCI_CGAP_119"
 /lab_host="DH10H (T1 phage-resistant)"
 /note="Organ: liver; Vector: pCMV-Sport6; Site 1: Not 1;
 Site 2: SalI; Cloned unidirectionally. Primers: oligo dT
 Average insert size 1.6 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 149..216..229..348..others

HASH COUNT 149..216..229..348..others

ORIGIN

Query Match 26.5%; Score 475.2; DB 10; Length 754;
 Best Local Similarity 78.6%; Pred. No. 1e 96;
 Matches 592; Conservative 0; Mismatches 159; Indels 2; Gaps 2;

QY 76 ctctctcagcag 134
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 2 CTCTCTTAT 61

QY 135 tcaagaggttggagcctcgggtcccaagagcctgaaaggtccacacacacacacacagagag 194
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 62 TCACAGCCGCTGAACACCACTAAGAGGCTGCTGAGCAAGCAAGCAAGCAAGCAAGCAAGCA 121

QY 195 ggccttgcctcag 254
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 122 GGCCTTCTATGCGGCACTGCTGACCTGCTGCGGCAACACACACACACACACACACACAC 181

QY 255 tgagcagagcctacggggagcctcctgagagagagagagagagagagagagagagagagag 314
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 182 TGATGAT 241

QY 315 tgaagcagcctgagac 374
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 242 TGAGCGGCTTGAACACCACTAAGAGGCTGCTGAGCAAGCAAGCAAGCAAGCAAGCAAG 301

QY 375 ggcctgacccctac 434
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 302 GACCAGACCTTACAGCTTACACTTATATATATATATATATATATATATATATATATAT 461

QY 435 actctgagcaggttgaggtgagcctcctgagagagagagagagagagagagagagagag 494
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 362 ACTCTGATACCTGCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 421

QY 495 ccatgctcctgacacagcctcctcctcctcctcctcctcctcctcctcctcctcctcct 554
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 422 CCATAGCTTCCGACCGCAGCTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 481

QY 555 agactaagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 614
 || || || || || || || || || || || || || || || || || || || || || ||
 Db 482 AGGCTAACCATCTCTGTCAGCAAGCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 541

cDNA clone 1300003707 5', mRNA sequence.

```

ACCESSION      BB610038
VERSION        BB610038.1
KEYWORDS       GI-16451655
SOURCE         EST.
ORGANISM       house mouse.
MUS MUSCULUS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 988)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tadawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suhiro cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-4222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.sc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subcloning of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes Genome Res. 10 (10), 1617-1630 (2000)
Wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Mutsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaoka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.sc.riken.go.jp) for
further details.
c mouse tissues.
FEATURES
    source
        location/Qualifiers
            1..988
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone_lib="RIKEN full-length enriched, adult male liver"
                /sex="male"
                /tissue_type="liver"
                /dev_stage="adult"
BASE COUNT    225 a 282 c 266 g 213 t 2 others
ORIGIN
Query Match      26.4%; Score 471 B; DR 9; Length 988;
Best Local Similarity 77.2%; Pred No. 6 1e-96;
Matches 611; Conservative 0; Mismatches 174; Indels 6; Gaps 3;
QY 9 acaacccgtgcacatcctcagcagcctcctcctacagttgtacagatggcagctgtccacgt 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68 ACAACCTGGAAATCTTAAGAGGCTGGCTTT--ATGATGTAATGATGATTTTCTTCTAGT 125

```

```

QY 69 ctattcccttcttcagccacacagaggttcttctcagcctctcagcctcttctcagcctcttct 128
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 ACAT--GCTCCCTTAGCCCTAGCCAGAGCTGCTACTAGGCTACAGGCTATCTTCTCTTCTTCT 182
QY 129 ggaatgctcaagaggttcttcagcctctcaggtctcccaagcctctcaaaatctccacacacacat 188
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 GATATGTTAAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTA 242
QY 189 ggaatgagccttctcagcagcctctcagcagcctctcagcagcctctcagcagcctctcagcagc 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 GGGGATTTGGGATTTATTTGGGATTTATTTGGGATTTATTTGGGATTTATTTGGGATTTATTTGG 402
QY 249 caaagatgaacacacacacacacacacacacacacacacacacacacacacacacacacacacac 408
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 CACGGCTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 462
QY 369 tgaatgctcagcagcctctcagcagcctctcagcagcctctcagcagcctctcagcagcctctcag 468
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 TGGTGTAAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTA 422
QY 369 agggcagcctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcag 428
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 AGGAGTAAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTA 482
QY 429 gcaacacactcagcagcctctcagcagcctctcagcagcctctcagcagcctctcagcagcctctcag 488
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 ACCACACATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 542
QY 489 ccttctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcagcctcag 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 GTTTTAAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 602
QY 549 gcaagagagctcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcag 608
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 GGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 662
QY 609 tcaacacctcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcag 668
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 663 TCAACACCATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 722
QY 669 gcaagacactcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcag 728
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 GGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 782
QY 729 tgaagagagctcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcagcctcagcag 788
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 783 CTGTGAGAGA-TGTCATCTCAAGGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 841
QY 789 ctacacctcagc 799
||| |||
Db 842 TCACCCCGGCC 852

```

```

RESULT      9
BF383882
LOCUS       602045401F1 NC1_CCAPI_L19 Mus musculus cDNA clone IMAGE:4194697 5',
mRNA sequence.
ACCESSION   BF383882
VERSION     BF383882.1
KEYWORDS    GI:11365187
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 860)
AUTHORS    NIH-MDC http://mac.ncbi.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, ph.D.
            Email: cqaabp@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.

```



```

Db 245 CACGGCTGACGACAGCACTATGGGACAGGCTGCTGACATGCGCATGCTGCTGG 304
QY 309 Lqqlqclqaacgcgcclqgacacacacacacacacacacacacacacacacacac 368
Db 305 TGGTGGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 464
QY 369 aggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 365 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
QY 429 qracacacacacacacacacacacacacacacacacacacacacacacacacac 488
Db 425 ACCCAGACACTGACACCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 484
QY 489 ccttcctccacacacacacacacacacacacacacacacacacacacacacacac 548
Db 485 GCTTCTCTCATAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
QY 549 gcaagggagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 608
Db 545 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
QY 609 lqacacacacacacacacacacacacacacacacacacacacacacacacacac 668
Db 605 TCGAACACAGTCACCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 664
QY 669 qraq-qacitccctcaagagagagagagagagagagagagagagagagagagag 727
Db 665 GGAAGAAATTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 724
QY 728 ttcgtgagagactgcctcctcctcctcctcctcctcctcctcctcctcctcctcct 787
Db 725 TTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784
QY 788 cctaac 793
Db 785 CCAACC 790

RESULT 11
BI217634
LOCUS 746 bp mRNA linear EST 11-Jul-2001
DEFINITION 602934072F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5097472 5',
mRNA sequence.
ACCESSION BI217634
VERSION BI217634.1 GI:14671078
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1
AUTHORS NIH-MGC http://mgc.nhl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.G.E. Consortium (I.M.N.I)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1233 row: p column: 17
High quality sequence stop: 746.
Location/Qualifiers
1..746
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5097472"
/clone_lib="NCI_CGAP_L19"

```

FEATURES

```

RESULT 12
BI328979
LOCUS 868 bp mRNA linear EST 30-Jul-2001
DEFINITION 60294034RF1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5134538 5',
mRNA sequence.
ACCESSION BI328979
VERSION BI328979.1 GI:15013636

```

```

/label="pH10B (T1 phage resistant)"
/Note="Organ: liver; Vector: pCMV-Sport6; Site: 1; Not 1;
Site: 2; SalI; Cloned unidirectionally. Primer: 6140 dir.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 164 a 221 c 209 g 152 t
ORIGIN
Query Match 26.2%; Score 469.2; DB 10; Length 746;
Best Local Similarity 78.8%; Pred. No. 2; 40-95;
Matches 586; Conservative 0; Mismatches 15; Indels 5; Gaps 2;
QY 9 acacacacacacacacacacacacacacacacacacacacacacacacacacac 68
Db 8 ACACACACACACACACACACACACACACACACACACACACACACACACACACAC 65
QY 69 ctgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 128
Db 66 ACAT---CTCTTAGCCCCAGAGCTGCTACTGCCACATGCCATCTCTGCTTACTG 122
QY 129 qgqlqclcaaaqqltllqagqccclqgqclcccccacacacacacacacacacacac 188
Db 123 GGATGCTCAGACGCTCAAGCAGCCAGCTTCCCAAGAGGCTCAAGAAATCCACGCC 182
QY 189 ggggtggcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 248
Db 183 GGGGCTTGGCTTTCATTGGGCAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 242
QY 249 caaggtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 308
Db 243 CACGGCTGACGACAGCACTATGGGACAGGCTGCTGACATGCGCATGCTGCTGG 402
QY 309 tggqlqclqaacgcgcclqgacacacacacacacacacacacacacacacacacac 368
Db 303 TGGTGGCTGACGCGCTGAAACACCATCAAGCAGCGCTGCTGACGAGGAGATGCTTCA 462
QY 369 aggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 363 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
QY 429 gcaacacacacacacacacacacacacacacacacacacacacacacacacacac 488
Db 423 ACCCAGACACTGACACCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 482
QY 489 ccttcctccacacacacacacacacacacacacacacacacacacacacacacac 548
Db 483 GCCTTCCCATAGCTCGGACCGCCACGCTCAGCAATCCCTGCTTATTTGAGGAGCA 542
QY 549 gcaagggagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 608
Db 543 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
QY 609 tcaacacacacacacacacacacacacacacacacacacacacacacacacacac 668
Db 603 TCAACACACACACACACACACACACACACACACACACACACACACACACACACAC 662
QY 669 qacacacacacacacacacacacacacacacacacacacacacacacacacac 728
Db 663 GCAACACACACACACACACACACACACACACACACACACACACACACACACACAC 722
QY 729 tgggtgagagactgcctcctcctcctcctcctcctcctcctcctcctcctcctc 752
Db 723 TTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 746

```

```

KEYWORDS      EST.
SOURCE         house mouse.
ORGANISM       Mus musculus.

REFERENCE
AUTHORS       Mammalia, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi:
TITLE         Eukaryota, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae; Mus
JOURNAL       1 (bases 1 to 868)
COMMENT       NIH-MGC http://mgc.nci.nih.gov/.
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@rs-remail.nih.gov
              Tissue procurement: Jeffrey E. Green, M.D.
              cDNA library Preparation: Life Technologies, Inc.
              DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM1141 row: 6 column: 11
              High quality sequence start: 4
              High quality sequence stop: 793.

FEATURES
source
1..868
Location/Qualifiers
1..868
/organism="Mus musculus"
/strain="FVB/N"
/clone="IMAGE:5133538"
/clone_lib="NCL-119"
/lab_host="FHL08 (T1 phage-resistant)"
/notes="Organ: liver, Vector: pCMV-SPOK16, Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL-119 library"

BASE COUNT    193 a 253 c 241 g 181 t
ORIGIN

Query Match      26.2%; Score 459.2; DB 10; Length 868;
Best Local Similarity 74.1%; Pred. No. 2 4e-95;
Matches 614; Conservative 0; Mismatches 163; Indels 7; Gaps 4;

QY 11 aacctgccaatctcaagcaccctgctctacagttggtacagatggcattgtccagctg 70
Db 11 AGCTTCGACAGCTCCTACGACCTCCGACCTGCTACAGATGCGCTTCCTCCAGTAC 69

QY 71 gtlccctctcagccacagagatctctcgtcctcgtcctcgtcctcgtcctcgtcct 130
Db 70 AT---CTCCTTAGCCCGACAGAGCTGCTACCTGCGCACCTGCGCACCTGCTGCTG 126

QY 131 gtctcaagggtttgagcctcggttcccaagagcctgaaagtccacacagcctgg 190
Db 127 ATGTCAGAGCCTCAAGAGAGTAAAGTTCCTAAAGAGCTGAGAAATCTACCTGAG 186

QY 191 ggtggtcctctcgtcgtgctgacccctggggaagaaacccgacacctggcactgtca 250
Db 187 GGCTTCCTCTTATCGGCACATGCTGACCTGCGGCAAGACCCACACCTGCTACAGCA 246

QY 251 aqataaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 310
Db 247 CGGTCAGTACAGCAGTATGGGACAGTGGCTGACATGCGGACAGTGGCTGCTGCTG 306

QY 311 gtctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 370
Db 307 GTGTTAGAGAGCTGAAATACATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 366

QY 371 ggcggcctgacctctacacctccacacacacacacacacacacacacacacacacac 430
Db 367 GCGCGACACACCTCTACAGCTTCACACTTATACATACAGCGCAAGAGGATGCTTCAAC 426

QY 431 acagactctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 490
Db 427 CCAGACTCTGGACCGCTGTGGGCTGGCCCGCGCGCTGGCCCGCGCTGGCCCGCTGA 486

```



```

QY 929 gagaagattgcaacctgtcaatgacatctttggagcaggatttgacacagtcaccaca 988
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 GAGAAGATTGCAACATTGTCATGACATCTTTGGAGCTGGCTTTTGAACACTACCA 600
QY 989 gccatctctcgg-agcctcatgtacattgtgaccagcctgagatgacaggaagatcca 1047
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 GCCATCACTGGAGCAATTTGCTACTGTTGACATGGGCTAAAGTGGCAGCAAGATCCA 660
QY 1048 qaaggaactgagacactgcatgagcaggagcggcggccggc-gctctcagacagccccc 1106
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 TGAGGAGCTGCAACAGCTGGCTGGGAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 720
QY 1107 agctgcccact-tggaggccttcactcctggagaccttcctcgacacactcctcttctggcc 1165
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 AGCTGCTATATATATATATATATATATATATATATATATATATATATATATATAT 780
QY 1166 ----ttaccatccccacagcacaacaggagacacacagctg-aatggcttctacatccc 1221
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CTTTCAACAATCCCCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
QY 1222 caagaaatgctgctct-ctglaaacacagtcgagctgacacatgacccag-agctgctgag 1279
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 CAAGGAGCCCTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 900
QY 1280 gaggagccctctdagtctcggcctgagcggtt 1311
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 GAAAGACCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 932

RESULT 15
BI144965
LOCUS 602909618F1 NCI_GCAP_Li9 Mus musculus cDNA clone IMAGE:5050685 5',
DEFINITION mRNA sequence.
ACCESSION BI144965
VERSION 1 (bases 1 to 835)
KEYWORDS EST.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaaps@mail.nih.gov
Tissue procurement: Jeffrey E. Green, M.D.
CDNA library preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11138 row: c column: 06
High quality sequence stop: 829
Location/Qualifiers
1..835
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5050685"
/clone_lib="NCI_GCAP_Li9"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_GCAP Library."
BASE COUNT 178 a 255 c 231 g 171 t
ORIGIN

```

```

Query Match 25.6% Score 458.8, DB 10, Length 835;
Best local Similarity 78.2%; Pred. No. 5,1e-93;
Matches 640; Conserved 0; Mismatches 167; Indels 11; Gaps 7;
QY 9 acacccctgccaatctcaagacacccctgacctctacagttgagacagatgacattgacaaat 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21 ACAACCTGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 ctgtctcctctcggccacagagctctctcctgacctctgacctctctcctgacctctctct 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79 AATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 135
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 ggggtctcaggggtttgagcctcgggtcccccagggcgtgaaaaggtccacacagacat 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 GGAGAGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 195
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 gggagctgacctctgctcggagcagtgctcaccctgggagagagagagagagagagagat 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 GGGGCTTTGGGCTTTGGGCTTTGGGCTTTGGGCTTTGGGCTTTGGGCTTTGGGCTTTGG 255
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 caaggaatgagacacagctctcggagagagagagagagagagagagagagagagagag 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 CACGAGTGAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTA 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 ctgggtctgagcctgagacacacacacacacacacacacacacacacacacacacacac 367
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 GTGCTGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 aagggcggcctcactctacacacacacacacacacacacacacacacacacacacacac 427
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 376 AAGGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 agcacagctctctgagcgggttgtagggctgcccggcggcggcggcggcggcggcggcgg 487
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 AAGGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTA 495
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 acctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 547
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 AATCTGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 555
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 agcacagagctctgagcgggttgtagggctgcccggcggcggcggcggcggcggcggcgg 607
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 AGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 615
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 608 ttgagcctctacacacagctctgagcgggttgtagggctgcccggcggcggcggcggcgg 667
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 616 TTGAAATGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTA 675
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 668 ggacagcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 726
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 GGAATAAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 735
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 727 gttcctgagagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 784
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 CTGCTGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTA 795
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 785 ctgacctacacacacacacacacacacacacacacacacacacacacacacacacacac 822
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 TGGAGTGAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 831
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: September 8, 2002, 16:54:42
Job time: 7296 sec


```

DB 71 MSKRFVAVVQIQMPGVVWVIRSFTHQAKTQKQVWVAGZGFLVSLVLDREKELAPSS 190
QY 128 DSHVWAAREELACATNATFETASTFASSECYTHLHVFKAKALRGLRGLHMAFCHDQ 187
DB 131 11111111111111111111111111111111111111111111111111111 111
DB 141 DQAVVWRTREKLAMALPSESTIKSCYSCVLLHVVKECHYVVKLSVNTASNSFD 190
QY 188 PYRVVSVANVIGAMTQGHFTUSDEMELSAVKHHEHVEHANSQFDEFTTELETPH 247
DB 191 11111111111111111111111111111111111111111111111111111 111
DB 191 PERHVSVANVTCGMEFGKRNHDDQELLSVNLSDREKQVAGSGNADEFTFLRPLPS 250
QY 248 PALQFFKAFNCRFLWELRIVVPHYQYQFQPS-----KQPSACNMLPCH 293
DB 251 KMFPTGTHARHVAVTVLHVEFTFENRSTCTGLLGHDEDELEKSNVQVWSD 310
QY 294 FVHVVFVAGTPTVATLSWLMVAVTEFEIQSLQKLLVIGRPRPRISDRQL 353
DB 311 11111111111111111111111111111111111111111111111111111 111
DB 311 KIVLVNGLFGAGFTLSALSWAVVLYVAPQKALQKQKQSVGLRPLLSKANL 370
QY 354 PYLEAFITLFRSSSLPFTPHISTRTOTTINQYIPKQCVFVQWGVNHHDPHWDPS 413
DB 371 PLEAFITLFRSSSLPFTPHICTKTALDQYTPKQFVQVFNQWGVNHHDPHWDPS 430
QY 414 EEPERFLTAISDAINPFLSEKMLFQZKQKPSCTGVVAKWHTFLFLALLOGLFVWP 473
DB 441 SPSEPTFLASVAVVSPKAFVWVMSKPEKSGHAGTFTVPLELALVGLFTFLP 490
QY 474 GVKVDLPIYGLMKHARCHEVQAKLR 500
DB 491 GHPLGIPYGLMKHKRQVLEASLR 516

RESULT 12
QYF1Y6 PRELIMINARY: PRI: 519 AA.
AC QYF1Y6
DT 01-MAY-2000 (TREMLOL 13, Created)
DT 01-MAY-2000 (TREMLOL 13, Last sequence update)
DT 01-DEC-2001 (TREMLOL 19, Last annotation update)
DE CYTOCHROME P4501A1.
GN CYP1A1.
OS Anguilla japonica (Japanese eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguilloidei;
OC Anguillidae; Anguilla.
OX NCBI_TaxID:7947;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: LIVER;
RX PubMed 10489414;
RA Aoki T, Iwamura T, Kato H, Sato M.;
KT "Isolation and sequence analysis of the bel cytochrome P450 CYP1A1
gene";
RL Mar. Biotechnol. 1:371-375(1999).
NC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DB EMBL: AB015744; BAAR248.1;
DB BSSP: p00179; 1D16.
DB InterPro: IPR001128; CYP_P450.
DB Pfam: PF00067; p450; 1.
DB PRINTS: PR00385; p450.
DB PROSITE: PS00086; CYTOCHROME_P450; UNKRNWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 519 AA: 5844 RW: 3ED4AGDQ4R99416A CRC64;

```

Query Match 53.2% Score 1405.5; DB 13; Length 519;
 Best local similarity 52.9%; Pred. No. 13; 187;
 Matches 268; Conservative 83; Mismatches 141; Indels 15; Gaps 2;

```

QY 8 PPSAELLASATGCVFVWVIRLHVPRQGLKSTPTWVWVGLGVHLLKRNHRAER 67
DB 11 FVSVSLAMAVCVVLMLELLEFTTEDEKPLKFLQVNVVWVWVWVWVWVWVWVWVWV 70
QY 69 MSQFVNAVIGTQSHVIVLSKLTIRCAVQKQIRKREHNTSTLTLLHLSLSTPT 127

```

```

DB 71 MSKRFVAVVQIQMPGVVWVIRSFTHQAKTQKQVWVAGZGFLVSLVLDREKELAPSS 190
QY 128 DSHVWAAREELACATNATFETASTFASSECYTHLHVFKAKALRGLRGLHMAFCHDQ 187
DB 131 11111111111111111111111111111111111111111111111111111 111
DB 141 DQAVVWRTREKLAMALPSESTIKSCYSCVLLHVVKECHYVVKLSVNTASNSFD 190
QY 188 PYRVVSVANVIGAMTQGHFTUSDEMELSAVKHHEHVEHANSQFDEFTTELETPH 247
DB 191 11111111111111111111111111111111111111111111111111111 111
DB 191 PERHVSVANVTCGMEFGKRNHDDQELLSVNLSDREKQVAGSGNADEFTFLRPLPS 250
QY 248 PALQFFKAFNCRFLWELRIVVPHYQYQFQPS-----KQPSACNMLPCH 293
DB 251 KMFPTGTHARHVAVTVLHVEFTFENRSTCTGLLGHDEDELEKSNVQVWSD 310
QY 294 FVHVVFVAGTPTVATLSWLMVAVTEFEIQSLQKLLVIGRPRPRISDRQL 353
DB 311 11111111111111111111111111111111111111111111111111111 111
DB 311 KIVLVNGLFGAGFTLSALSWAVVLYVAPQKALQKQKQSVGLRPLLSKANL 370
QY 354 PYLEAFITLFRSSSLPFTPHISTRTOTTINQYIPKQCVFVQWGVNHHDPHWDPS 413
DB 371 PLEAFITLFRSSSLPFTPHICTKTALDQYTPKQFVQVFNQWGVNHHDPHWDPS 430
QY 414 EEPERFLTAISDAINPFLSEKMLFQZKQKPSCTGVVAKWHTFLFLALLOGLFVWP 473
DB 441 SPSEPTFLASVAVVSPKAFVWVMSKPEKSGHAGTFTVPLELALVGLFTFLP 490
QY 474 GVKVDLPIYGLMKHARCHEVQAKLR 500
DB 491 GHPLGIPYGLMKHKRQVLEASLR 516

RESULT 13
QYF1Y6 PRELIMINARY: PRI: 521 AA.
AC QYF1Y6
DT 01-MAY-1999 (TREMLOL 10, Created)
DT 01-MAY-1999 (TREMLOL 10, Last sequence update)
DT 01-DEC-2001 (TREMLOL 19, Last annotation update)
DE CYP1A.
GN CYP1A.
OS Fundulus heteroclitus (killifish) (Mammichthys).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Fundulidae; Fundulus;
OC Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;
OC Cyprinodontiformes; Fundulidae; Fundulus.
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: LIVER;
RX MEDLINE 99138412; PubMed 9972465;
RA Mortenson H, G. Wolf E, J. Katchen S, L. Sodin M, L. Stodman L, L.
KT "Molecular cloning of CYP1A from the estuarine fish Fundulus
heteroclitus and phylogenetic analysis of CYP1 genes: update with new
sequences";
RL Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol.
RL 121:241-240(1998).
NC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DB EMBL: AF026800; AA001809.1;
DB BSSP: p00179; 1D16.
DB InterPro: IPR001128; CYP_P450.
DB Pfam: PF00067; p450; 1.
DB PRINTS: PR00385; p450.
DB PROSITE: PS00086; CYTOCHROME_P450; UNKN WN 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 521 AA: 5990 RW: 393524152452A CRC64;

```

Query Match 53.2% Score 1405.5; DB 13; Length 521;
 Best local similarity 52.9%; Pred. No. 13; 187;
 Matches 268; Conservative 83; Mismatches 141; Indels 15; Gaps 2;

```

QY 1 MALQGVPR 645ELLSAVALTGVVWVIRKQKQIRKREHNTSTLTLLHLSLSTPT 127

```





